NRL Report 8678

Invariant Clustering Using Scattering Matrices

G. V. TRUNK

Radar Analysis Branch
Radar Division

February 23, 1983

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NRL Report 8678	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) INVARIANT CLUSTERING USING SCATTERING MATRICES		5. TYPE OF REPORT & PERIOD COVERED Final report on one phase of an NRL problem 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G.V. Trunk		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N RR021-05-43 NRL Problem 53-0690-0
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Navy Office of Naval Research Arlington, VA 22217		12. REPORT DATE February 23, 1983 13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Clustering
Scattering matrices
Invariant clustering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

One of the most difficult problems in clustering analysis is the choice of an appropriate distance metric. This problem arises because in many instances an object is described by n measurements involving different units (for example, a radar signal may be described by its frequency in gigahertz, its interpulse period in milliseconds, and its pulsewidth in microseconds). In this report I circumvent the problems by considering clustering algorithms which are invariant to scaling of the axes and hence to the choice of units. I compare quality of the clustering produced by the various algorithms by examin-

(Continues)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

ing several test cases. No single algorithm yielded the correct solution in all of the cases, but a possible hybrid approach is to cluster the points using the product of the determinants of the scattering matrices and switch to the sum of the determinants if the product algorithm yields only degenerate solutions.

CONTENTS

INVARIANT CLUSTERING USING SCATTERING MATRICES

INTRODUCTION

When dealing with practical clustering problems, one of the most difficult choices is the choice of the distance metric. This problem exists because if an object is described by n measurements involving different units, the Euclidean metric normally used has little or no meaning. For instance, electronic-warfare-support-measure (ESM) equipment may characterize a radar signal by its frequency in gigahertz, its interpulse period in milliseconds, and its pulse width in microseconds. Given signals from two agile (frequency, PRF, and pulse width) radars, a natural question is what signals are being generated by what radars.

All clustering algorithms tend to associate together points which are similar, similar points being close to one another in feature space. However, if the various features are measured in different units, it is extremely difficult to select the metric which will yield a "correct" clustering of the points. Thus, in this report I will consider an alternate approach involving clustering algorithms which are invariant to scaling of the axes and hence to the choice of units. I will compare the quality of the clustering produced by the various algorithms by using several test cases.

CLUSTERING ALGORITHMS

Duda and Hart [1] describe several algorithms which are invariant to nonsingular linear transformations: $y = A\,x$, where y and x are d vectors and A is a d-by-d matrix which has a nonzero determinant. These algorithms find a minimum of the determinant of the within-class scattering matrix S_w and a maximum of the trace $S_W^{-1}\,S_B$, where S_B is the between-class scattering matrix:

$$\text{minimize } |S_W|$$

and

$$\text{maximize tr } \left\{S_W^{-1}\,S_B\right\} = \sum_{i=}^{d} \lambda_i\,,$$

where d is the dimensionality of the observation space and $\{\lambda_i\}$ are the eigenvalues of $S_W^{-1}\,S_B$. The scattering matrices are defined by the following relationships:

- The mean vector for the ith cluster is

$$\mu_i = \frac{1}{n_i} \sum_{x \epsilon X_i} x,$$

where n_i is the number of points in the ith cluster and X_i is the collection of points in the ith cluster;

Manuscript approved December 6, 1982.

- The total mean vector is

$$\mu = \frac{1}{n} \sum_{X} x = \frac{1}{n} \sum_{i=1}^{c} n_i \mu_i,$$

where c is the number of clusters;

- The scatter matrix for the ith cluster is

$$S_i = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^t;$$

- The within-cluster scatter matrix is

$$S_W = \sum_{i=1}^{c} S_i;$$

- The between-cluster scatter matrix is

$$S_B = \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^t;$$

- The total scatter matrix is

$$S_T = \sum_{x \in X} (x - \mu)(x - \mu)^t.$$

It follows from these definitions that the total scatter matrix is the sum of the within-cluster scatter matrix and the between-cluster scatter matrix:

$$S_T = S_W + S_B.$$

Using this last equation, one can derive two other invariant criterion functions:

$$\mathrm{tr}\left\{S_T^{-1} S_W\right\} = \sum_{i=1}^{d} \frac{1}{1 + \lambda_i}$$

and

$$\frac{|S_W|}{|S_T|} = \prod_{i=1}^{d} \frac{1}{1 + \lambda_i}.$$

When there are only two clusters (one nonzero eigenvalue), the last three criteria will yield the same partitioning. However, when there are more than two clusters, the partitionings may not be the same.

My first attempts at investigating invariant clustering algorithms have been limited to minimizing the determinant of the within-class scattering matrix and variations of it.

DETERMINANT OF WITHIN-CLASS SCATTERING MATRIX

Since it is essentially impossible (too time consuming) to evaluate the criterion function for all possible partitions, I select an initial partition and then iteratively move points from one cluster to another until a minimum or maximum is obtained. Thus, assuming an initial partitioning, I will now state what happens to the various means and scattering matrices when a single point x^* is moved from the ith cluster to the jth cluster:

$$\mu_i^* = \mu_i - \frac{x^* - \mu_i}{n_i - 1} ,$$

$$\mu_j^* = \mu_j + \frac{x^* - \mu_i}{n_j + 1} ,$$

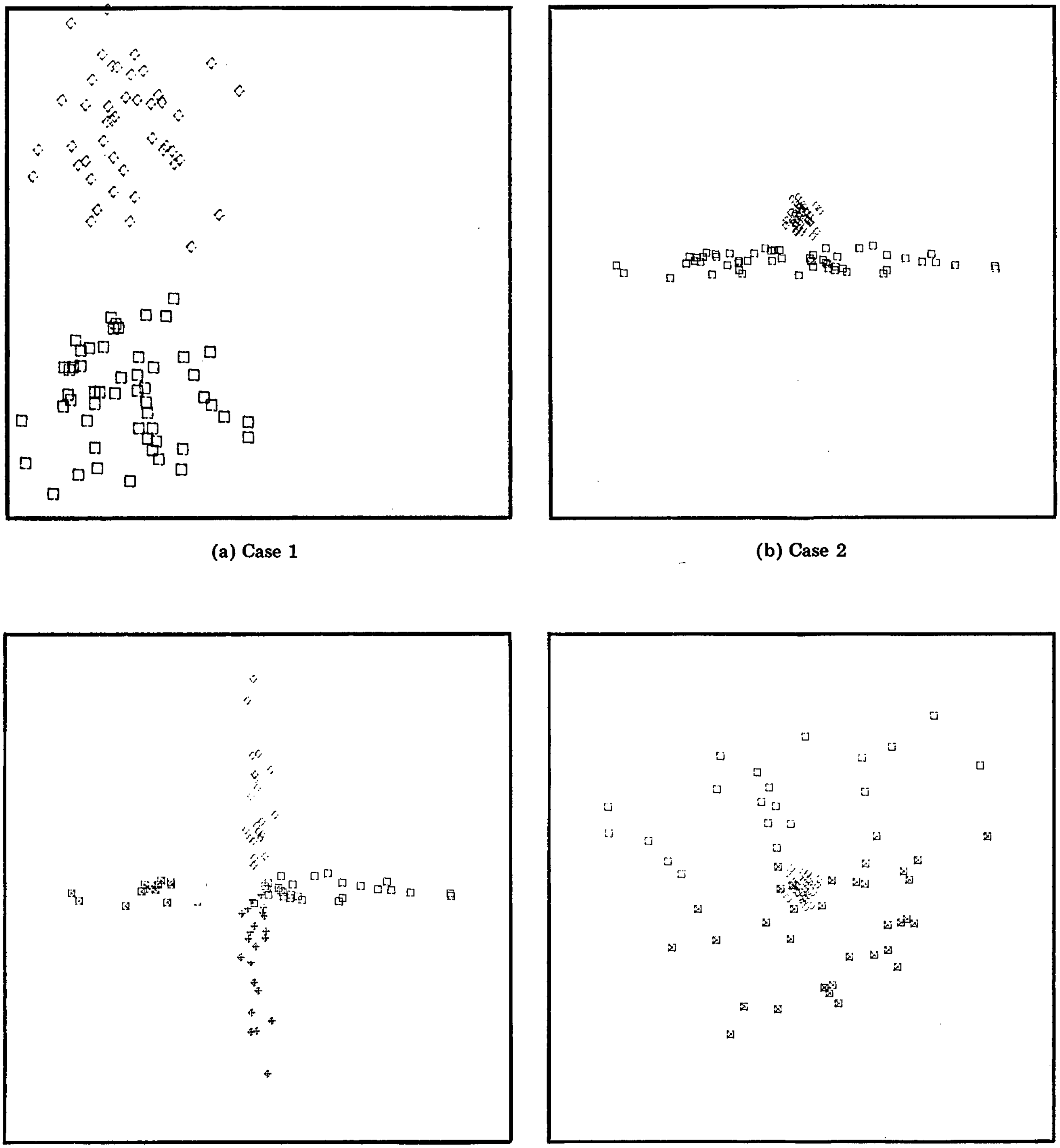
$$S_i^* = S_i - \frac{n_i}{n_i - 1} (x^* - \mu_i)(x^* - \mu_i)^t,$$

$$S_j^* = S_j + \frac{n_j}{n_j + 1} (x^* - \mu_j)(x^* - \mu_j)^t.$$

One should never remove x^* from the ith cluster if it is presently the only point in the cluster. Thus in my procedure one calculates the means, scattering matrix, and its determinant for an initial partitioning of the points. Next, one tentatively moves each point from its present cluster to each of the other clusters. If the determinant is smallest with x^* in the jth cluster, one assigns x^* to the jth cluster. One then iteratively cycles through all n points until no points move from one cluster to another. Of course such a procedure yields only a local minimum and not a global one. Therefore, in my study I used 25 different random starting positions.

To visualize the quality of the clustering, I investigated only two-dimensional, two-class problems. I chose the points in both classes from bivariate Gaussians with mean value M_i and covariance matrix K_i, with the means and covariances for my four cases being as shown in Table 1. In case 1, the clusters are well separated, and I would expect any clustering algorithm to yield the correct solution. Case 2 is more difficult, with a spherical clustering being fairly close to an elongated cluster. Case 3 is still more difficult, with the two clusters having the same mean but larger variances in orthogonal directions. Case 4 is the most difficult, with the cluster embedded within the other. I do not expect any clustering algorithm to yield the correct solution in this case, since I cannot envision a function for which the correct partitioning is an extremum point.

I minimized the determinant of the within-cluster scattering matrix iteratively, and the results for the four cases are shown in Fig. 1. The squares are from cluster 1, the diamonds are from cluster 2, and a cross in the square or diamond indicates that the point was classified incorrectly. In cases 1 and 2 the points were correctly classified, but in cases 3 and 4 the points were incorrectly classified. In the 25 repetitions of case 1, two local minimums were found, and the smallest which corresponds to the correct solution appeared 17 times. In case 2, three local minimums were found, and the smallest which corresponds to the correct solution appeared eight times. In case 3, two local minimums were found, but neither was associated with the correct solution, since the correct solution was not even a local minimum. Similarly in case 4, five local minimums were found, but none was associated with the correct solution, since the correct solution was not even a local minimum. As I



(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

Fig. 1 — Clustering results for the determinant of the sum of the scattering matrices

Table 1 — Parameters of Bivariate Gaussians

Case	Class 1		Class 2	
	Mean Vector	Covariance Matrix	Mean Vector	Covariance Matrix
1	0	1 0	0	1 0
	0	0 1	5	0 1
2	0	10 0	0	1 0
	0	0 1	5	0 1
3	0	10 0	0	1 0
	0	0 1	0	0 10
4	0	10 0	0	1 0
	0	0 10	0	0 1

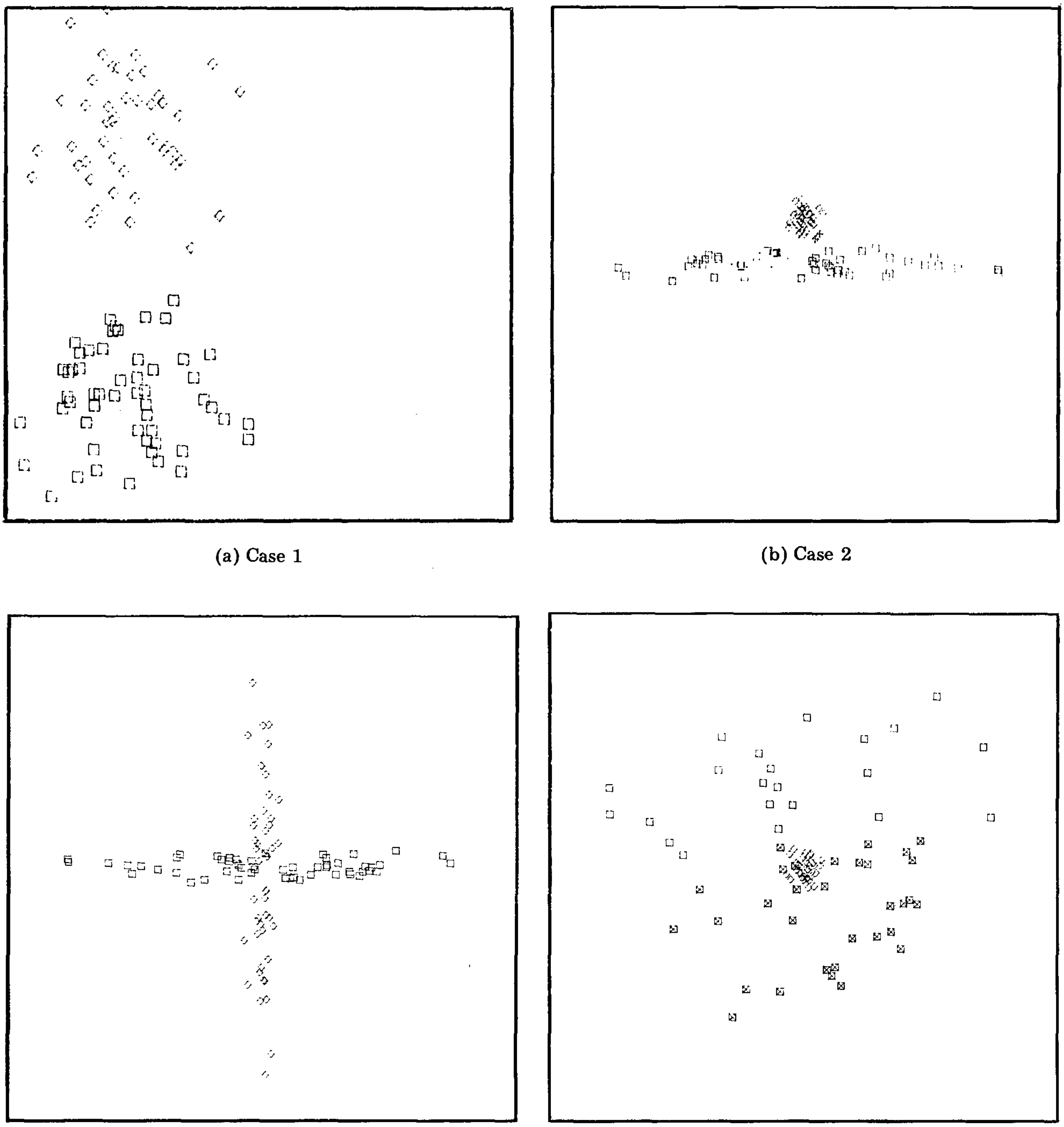
mentioned, I do not expect case 4 to be correctly handled; however, I do expect cases 1, 2, and 3 to be correctly handled. Consequently, I attempted to modify the criterion function.

Although the determinant of a scatter matrix in some sense represents a volume containing the points, it is not clear what geometric interpretation one can give to the determinant of the sum of scatter matrices. Thus, an obvious modification* is to minimize the sum of the determinants:

$$\min \sum_{i=1}^{c} |S_i| .$$

I again used iterative techniques to minimize the criterion function, and the results are given in Fig. 2. Cases 1 and 3 were correctly handled, but many points in cases 2 and 4 were incorrectly classified. In the 25 repetitions of case 1, three local minimums were found, and the smallest which corresponds to the correct solution appeared 17 times. In case 2, four local minimums were found, but none was associated with the correct solution, which was not even a local minimum. In case 3, two local minimums were found, and the smallest which corresponds to the correct solution appeared 15 times. Finally in case 4, five local minimums were found, none of which corresponded to the correct solution. Intuitively, the reason the wrong solution was obtained for class 2 was that the determinants of the individual scattering matrices were different by an order of magnitude.

*The proof that this algorithm is invariant to nonsingular linear transforms parallels the one showing min $|S_W|$ is invariant [1].



(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

Fig. 2 — Clustering result for the sum of the determinants of the scattering matrices

Thus, if one moves a few points from the large cluster to the small cluster, the determinant of the small cluster may double, but the determinant of the large cluster may decrease by 10%, still resulting in a smaller criterion function. To remedy this situation, I considered a criterion function which is the product of determinants.*

Specifically, I searched for the clustering which minimized

$$\prod_{i=1}^{c} |S_i|$$

and resulted in a nondegenerate solution. That is, if one puts all the points in one classs, the criterion function will have a value of zero. Therefore, one desires a clustering which corresponds to a local minimum of $\Pi\,|S_i|$ whose value is the smallest value greater than zero. Again I minimized the criterion function iteratively, and the results are given in Fig. 3. Although it appears that the clusters in cases 1, 2, and 3 were correctly identified, this is somewhat misleading. The problem is with case 1. The correct solution corresponds to the smallest nonzero local minimum, but all 25 random initial clusterings yielded degenerate solutions. In case 2, one nonzero local minimum was found, and the solution which corresponds to the correct solution appeared 12 times. The other 13 times a degenerate solution was obtained. In case 3, three nonzero local minimums were found, and the smallest which corresponds to the correct solution appeared eight times, the others just once. The other 15 times a degenerate solution was obtained. In case 4, all solutions were degenerate solutions—the correct solution is not a local minimum. Thus, the problem with this procedure apparently is that if clusters are "too" close to one another, there is a high probability of having a degenerate solution. Surprisingly the results do not improve quickly as the separations increase. Figure 4 represents case 1 with the separation in means increased from 5 to 10. However, although the correct solution is found five times, a degenerate solution is found 20 times. One possible way of proceeding is to cluster the points using $\Pi\,|S_i|$. If only degenerate solutions are found, one can switch then to either $|\Sigma S_i|$ or $\Sigma\,|S_i|$. Obviously, this procedure yields the correct solution to the three simple cases (1, 2, and 3).

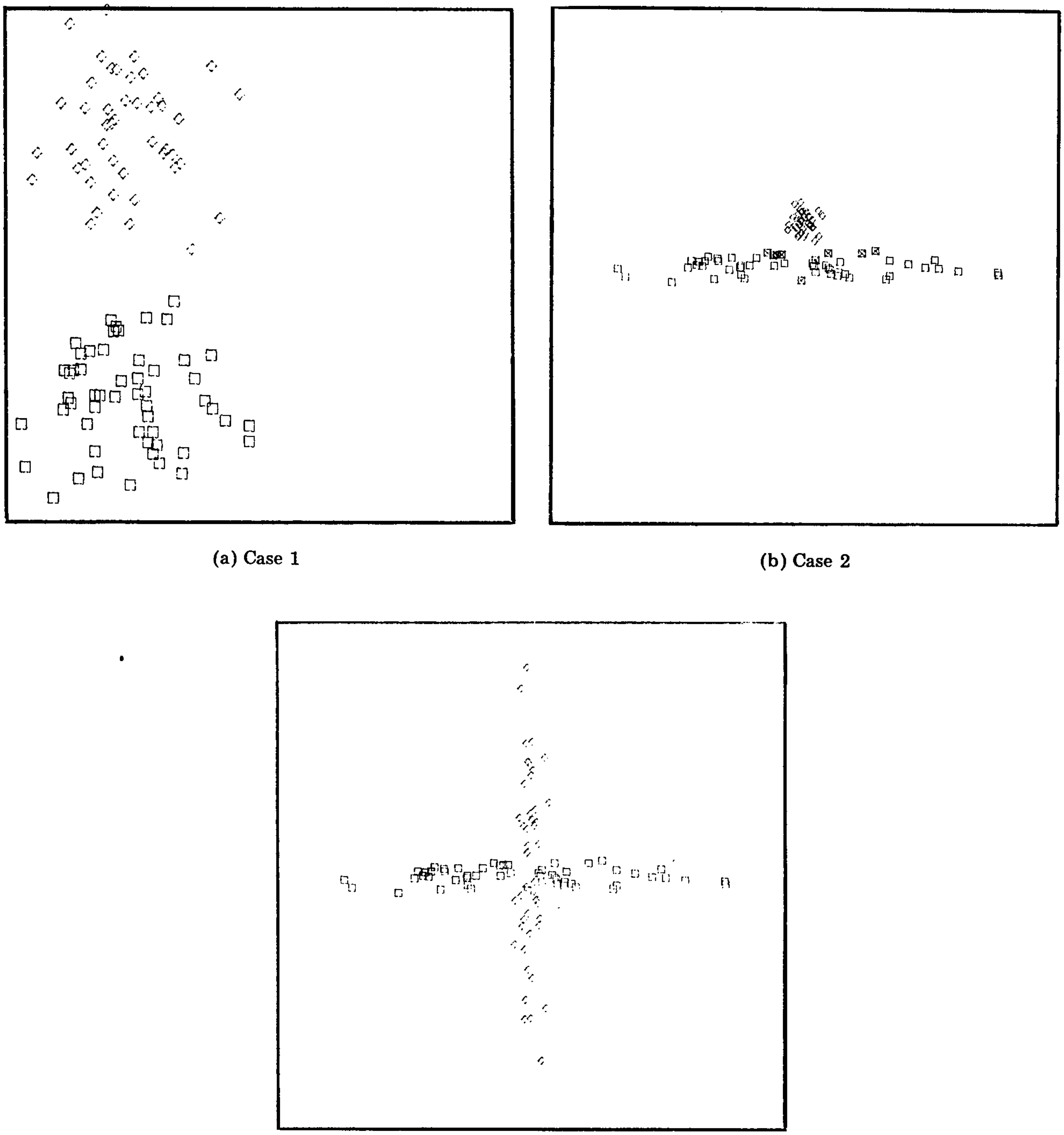
SUMMARY

Several invariant clustering algorithms based on scattering matrices have been investigated. No single algorithm yielded the correct solution to all of the three simple clustering cases in this report. One possible approach is to cluster the points using the product of the determinants of the scattering matrices and switch to the sum of the determinants if only degenerate solutions are found using the product algorithm. This hybrid method yields the correct solution to the three simple clustering cases.

REFERENCE

1. R.O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis," Wiley, New York, 1973.

*Again the proof that this algorithm is invariant to nonsingular linear transforms parallels the one showing min $|S_W|$ is invariant [1].



(a) Case 1

(b) Case 2

(c) Case 3

Fig. 3 — Clustering result for the product of determinants of the scattering matrices

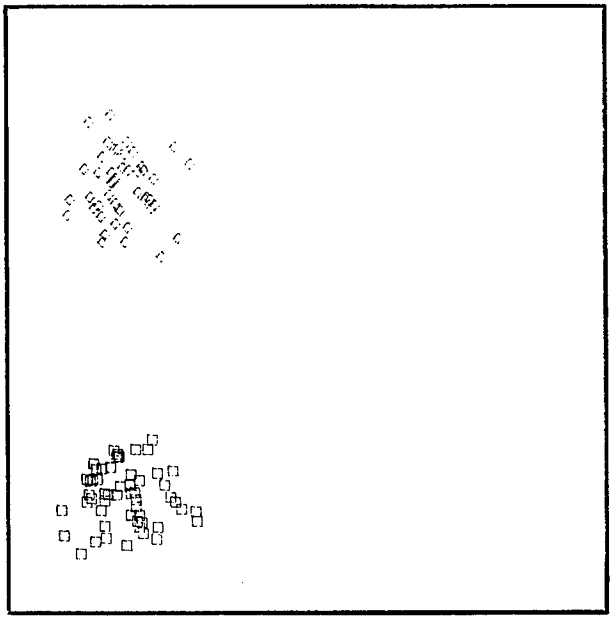


Fig. 4 — Clustering result for the product of determinants of the scattering matrices when case 1 is modified by increasing the separation of means from 5 to 10.